

## О ПАРАЛЛЕЛЬНОМ КОРПУСЕ РУССКОГО И КИТАЙСКОГО ЯЗЫКОВ В ЛИНГВОДИДАКТИКЕ

ВАН ЦИН  
БГУ, Минск

Сопоставление ключевых категорий системы языка актуально в лингводидактических и контрастивных исследованиях.

С течением времени значение информационных технологий непрерывно возрастает. Когда речь идет о создании перспективных информационных технологий, то на передний план выходят проблемы использования естественного языка в системах автоматической обработки текста. Такими проблемами занимается компьютерная лингвистика. Относительно новое направление, тесно связанное с компьютерной лингвистикой, называется корпусной лингвистикой. Корпусные исследования направлены на разработку общих принципов построения и использования лингвистических корпусов (корпусов текстов) с помощью компьютерных технологий.

Текст (устный и письменный) является одним из основных источников языкового материала. **Лингвистический корпус** – центральное понятие корпусной лингвистики – это коллекции текстов в электронном виде, размеченных по разным параметрам.

Использование корпусов текстов дает возможность эффективно исследовать языковые единицы, интересующие пользователя. Лингвистические корпусы представляются особенно актуальными при изучении иностранного языка.

В последнее время появились параллельные корпуса русского и китайского языков, один из которых – Текстовый корпус русского и китайского языков (далее – ТКРКЯ) функционирует по адресу: <http://rucorpus.cn> [3]. В него входит «корпус художественного стиля» (далее – КХС), «корпус публицистического стиля» (далее – КПС) и «обучающий корпус». КХС состоит из нескольких баз данных по авторам произведений. В данный момент доступна коллекция произведений А.П. Чехова, содержащая 257494 иероглифа. КПС состоит из трех баз данных: «РИА Новости», «Российские газеты» и «Первый канал России». Общее количество иероглифов – 3676. КПС и Обучающий корпус находятся в процессе создания.

В КХС по запросу слова или словосочетания можно получить конкорданс. **Конкордансом** называется список всех употреблений искомого слова в контексте со ссылками на источник. КХС в отличие от «Параллельного корпуса НКРЯ» дает не только варианты на русском и другом языках, но и характеристику предложения на соответствующем запросу языке, а также указывает степень сложности перевода.

Рассмотрим, например, лексему 生命 (shēng mìng – кит. ‘жизнь; жизнедеятельность’). Получаем следующий конкорданс:

1. 俄语: В самом деле, против его воли вызван он какими-то случайностями из небытия к жизни...

汉语: 确实, 他从不存在成为有生命的東西并非出于自己的意志, 而是由某种偶然的条件促成的,

语句属性: 陈述句, 简单句, 被动形动词, 俄汉句子数量不等, 俄汉句子结构不同, 译者加词处理, 俄汉词序不同, 原文拆分为多句;

语料库: 契诃夫短篇小说作品语料库.

2. 俄语: А органическая ткань, если она жизнеспособна, должна реагировать на всякое раздражение.

汉语: 人的机体组织, 如果是有生命的, 就必然对一切刺激有反应;

语句属性: 陈述句, 复合句, 形容词短尾, 主从复合句, 条件从句, 译者减词处理, 俄汉词序不同;

语料库: 契诃夫短篇小说作品语料库.

Таким образом, найдено два контекста со словом «生命». Характеристика первого предложения – повествовательное простое предложение с краткой формой страдательного причастия: *вызван*. При переводе трудность состоит в том, что структуры предложения на русском и китайском языках не одинаковы, с разным порядком слов. В данной ситуации переводчик добавил слова и разбил русский вариант на несколько фраз с самостоятельным значением. Можно открыть ссылку для того, чтобы увидеть более полный фрагмент текста.

Характеристика второго предложения – повествовательное сложноподчиненное предложение, содержащее краткую форму прилагательного: *жизнеспособна*, и обстоятельственным придаточным условия: *если она жизнеспособна*. При переводе трудность состоит в том, что структуры предложения на русском и китайском языках не одинаковы, с разным порядком слов, как в первой фразе.

Во втором примере используется не слово *生命*, а слово *有生命的* – ‘живой, одушевленный’. В русском тексте – *жизнеспособна*. Это слово состоит из трех фрагментов: 有 (*yǒu* – кит. обладать, иметь; у (кого-то) есть (что-то); быть владельцем), 生命 (*shēng mìng* – кит. ‘жизнь; жизнедеятельность’), 的 (*dě* – кит. синтаксическое служебное слово; формант слова, выступающего в функции определения и обозначающего). Затруднение при разделении слов возникает в связи с тем, что в китайском языке разница между существительным и прилагательным реализуется при помощи служебного слова *的*.

Имеются следующие рекомендации: «Для того, чтобы решить проблему, мы можем конкретизировать поисковое слово по ограниченным условиям, т. е. давать грамматическую характеристику, но эти условия только исключают часть бессмысленных слов» [4].

Для пользователей корпусная методология имеет важное значение: «Безусловно, лингвистический корпус имеет важное значение в изучении языка. Если лингвистический корпус двух и более языков построен

рационально, он обогатит метод исследования и изучения языков» [6, с. 192].

В лингвистическом корпусе можно быстро найти слова и конструкции с грамматическими и другими свойствами. В нем содержится ряд примеров разных жанров по различным временам. «Современные компьютерные программы могут учитывать индивидуальные потребности и интересы учащихся, различные стратегии овладения языком, дифференцировать способы предъявления учебного материала, обеспечивать индивидуальные формы тренировки, создавать широкий диапазон стимулов для вовлечения учащихся в иноязычную речевую деятельность, увеличивать время контакта с изучаемым языком» [2, с. 8].

Преподаватель с навыками использования корпуса может из пассивного пользователя словарей и грамматики превратиться в активного участника корпусной революции.

«Достижения в области корпусной лингвистики находят широкое применение в процессе преподавания языков» [5]. Корпусные данные в обучении языков становятся повседневным практическим материалом для лекционных курсов и самостоятельного выполнения заданий.

#### БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Большой китайско-русский словарь [Электронный ресурс]. Режим доступа: <http://bkrs.info>. (дата доступа: 23.05.2014).

2. Голубева Т.И. Применение информационных технологий в обучении иностранному языку: Учебное пособие / Т.И. Голубева, С.О. Репина. – Оренбург: ГОУ ОГУ, 2004. – 167 с.

3. Текстовый корпус русского и китайского языков [Электронный ресурс]. Режим доступа: <http://rucorpus.cn>. (дата доступа: 23.05.2014).

4. Lamy M-N., Klarskov-Mortensen H. J. Using concordance programs in the Modern Foreign Languages classroom [Электронный ресурс]. Режим доступа: [http://www.ict4lt.org/en/en\\_mod2-4.htm](http://www.ict4lt.org/en/en_mod2-4.htm). (дата доступа: 22.05.2014).

5. 刘守军 词类 (Категория слов) / 刘守军 [Электронный ресурс] Режим доступа: [http://study.hhit.edu.cn/subject/CourseWare\\_Detail.aspx?TeachCourseWareID=1684/](http://study.hhit.edu.cn/subject/CourseWare_Detail.aspx?TeachCourseWareID=1684/) (дата доступа: 04.05.2014).

6. 王建新, 计算机语料库的建设与应用 (Создание и применение лингвистического корпуса) / 王建新. – 清华大学出版社, 2005. – 311 с.

7. 在线新华字典 (Онлайн-словарь Синьхуа) [Электронный ресурс]. Режим доступа: <http://xh.5156edu.com>. (дата доступа: 26.05.2014).